
# Crossing between environments

Andrews, Tegau; Prys, Gruffudd; Prys, Delyth; Jones, Dewi

## Terminologie(s) et traduction

*DOI:*
10.3726/b14740

Published: 01/01/2018

Peer reviewed version

# Crossing between environments: the relationship between terminological dictionaries and Wikipedia

Tegau Andrews, Gruffudd Prys, Dewi Bryn Jones and Delyth Prys
Language Technologies Unit, Bangor University, techiaith@bangor.ac.uk

This paper discusses two environments in which concept definitions are published, namely terminological dictionaries which deal specifically with technical subjects and general online encyclopaedias such as Wikipedia. It compares the intended function of both resource types, justifies and explores their differences, and considers whether a mutually beneficial relationship can exist between both knowledge formats. This discussion will focus on a real world exercise undertaken to enrich term entries and definitions relating to the natural world in Welsh terminological dictionaries. The paper also investigates the technological methods used to incorporate Wiki content into terminological dictionaries, and examines some of the licensing issues arising from sharing content between different resources. It concludes by examining how the interconnection between the environments of Welsh Wicipedia and of Welsh terminological dictionaries might be developed in the future.

**Key words:** terminological dictionaries, terminological definitions, crowdsourcing, Welsh Wicipedia, Wikipedia, content sharing

This paper was given initially as a presentation at the Bucharest TermTrad 2017 International Symposium on *Terms of the environment and the environment of terms*. We chose to interpret the symposium title as referring to the environment (as in platform, or setting) in which terms and definitions are published and distributed. This paper was written in response to an increasing tendency in Wales, observed by the authors, to question the need for terminological dictionaries in light of the growth of collaborative online encyclopaedias. It compares the intended function of both types of resources, exploring and justifying the differences between them, and considers whether a mutually beneficial relationship can exist between both knowledge formats. It also examines these issues in the context of a real-world exercise to enrich term entries and definitions relating to the natural world within the field of Welsh-medium education in Wales.

Terminological dictionaries and crowdsourced online encyclopaedias are often seen as being similar resources. Both represent collections of entries that label and describe concepts for the benefit of their readers. However, it is the encyclopaedic format which is the most familiar generally, and the growth of one popular example, Wikipedia, has served to further eclipse the more specialized medium of the terminological dictionary.

The success of the crowdsourcing approach for Wikipedia and similar platforms has led to a discussion within the field of terminology as to whether there is a place for crowdsourcing within the discipline. This question has been raised regularly and repeatedly in terminology conferences, including in the *Creation, Harmonization and Application of Terminology Resources* NODALIDA workshop in 2011, the *Multilingual Terminology Development in Higher Education: Terminologies, Glossaries and the Academic* event at the University of South Africa in 2014, and the 2016 EAFT Summit in Luxembourg. We address the role of crowdsourcing below, and consider whether more open collaboration would be appropriate when standardizing terminology.

In the education sector in Wales, some have gone beyond suggesting that terminology work should imitate certain aspects of the approach adopted by Wikipedia. They now ask whether the existence of a Welsh Wicipedia has done away completely with the need to produce specialist terminological dictionaries any longer. In the September 2017 EAFT seminar in Bangor University, Wales, the following question was raised by a representative of a body responsible for funding Welsh-medium education:

> Do we need Welsh terminological dictionaries at all? Should we not input all of their contents into Wikipedia, and do away with terminological dictionaries? This would make Wikipedia a one-stop shop for students – it is, after all, so much more popular as a source of information for them.

Put simply, the question is would terminology work in Wales benefit from increased crowdsourcing, or, more drastically, should we adopt Wikipedia wholesale as a distribution outlet for Welsh-language terminology. To answer that we need to look carefully at these two distinct environments, compare and contrast what each has to offer and decide for whom they might work best.

Terminological dictionaries contain technical terms from a particular domain, along with other information such as a concise definition of the concept for which the term serves as a label. International best practice as set out by the International Organization for Standardization (ISO) suggests that terms should be standardized by a working group consisting of five to eight domain experts and a terminologist (British Standards Institution 2001: 4.3.4). In this environment candidate terms are collected and evaluated using ISO 704 principles, before consensus is reached between the working group members on the recommended standardized form. These principles include ensuring that the term is transparent, appropriate, concise and linguistically correct, and that it can give rise to derivatives and compounds (British Standards Institution 2009: 7.4.2).

Definitions are drafted and fine-tuned by the working group, who use a mixture of domain-expert knowledge, reference books and other such sources and follow the ISO 704 principles on definition writing. Among other things, these principles state that a definition should include the unique set of characteristics that typify the concept, enough information so that the reader can recognize the concept and differentiate between it and other related concepts, and be as concise as possible (British Standards Institution 2009: 6.2).

Terminological dictionaries are prescriptive works targeted at a particular audience, often specialists working in the relevant field. Prescriptive dictionaries are described as such because they *prescribe* the recommended term to be used in a specific technical context with a particular intended audience. They differ from the more common descriptive dictionaries where many different synonyms or alternative phrases are provided, leaving the choice of which word to use in the hands of the user. Unlike descriptive general-language dictionaries, technical prescriptive dictionaries are usually written by a small group of domain experts aided by linguists. As a result, creating such technical dictionaries could be described as a collaborative, multidisciplinary endeavour between a number of experts seeking to achieve consensus on the most appropriate definition and designation for the concepts to be included in the dictionary.

The Wikipedia environment, on the other hand, is very different. Wikipedia is a general-purpose resource intended for the widest possible audience. Wikipedias in major languages, such as English, are created and edited by a large number of anonymous volunteers. A Wikipedia article does not need to be written in a technical language register and in fact it could be argued that the need to be clear and "understandable […] for both experts and non-experts" ("Wikipedia: The Perfect Article" n.d.) means it should not be written in a technical register. According to the English Wikipedia guidelines, an article should be "long enough to provide sufficient information, depth and analysis on its subject" ("Wikipedia: The Perfect Article" n.d.) and should include references, citations and relevant media. This goes considerably beyond the scope of a term definition. A term definition could, however, be contained within the body of an expanded article. Wikipedia is freely available online and one of its most notable characteristics is that anyone may contribute. On Wikipedia itself, it states: "*anyone* can edit almost every page" ("Wikipedia: Introduction" n.d.). Therefore, one of the biggest differences between Wikipedia and terminological dictionaries is authorship – the individuals who are permitted to write its content. In one case it is a specific group of specially selected individuals. In the other it is an unknown number of self-selected individuals.

Wikipedia's open, crowdsourcing philosophy has resulted in its expansion to comprise 35 million articles in 288 different languages (Safer 2015). These articles are not, however, distributed equally between all languages. As a rule of thumb, the more contributors to Wikipedia a language has, the more content is created in that language. A crowdsourced approach succeeds when it attracts a critical mass of contributors and editors to expand and improve the resource for free. Attracting freely-given

contributions is easier when a language has more speakers. For example, Figure 1 shows the number of articles available in various European languages on 14 May 2018.

**Figure 1: Comparison of the number of articles for various languages**

| Articles | Language |
|---|---|
| 5,648,850 | English |
| 3,784,073 | Swedish |
| 1,983,143 | French |
| 1,410,459 | Spanish |
| 579,906 | Catalan |
| 430,983 | Hungarian |

Source: List of Wikipedias n.d.

Compare this with the situation for a language with fewer speakers, such as Welsh. Welsh is one of six Celtic languages, and it is spoken in Wales by 19% of the population, or approximately 562,000 speakers (Office for National Statistics 2012). It is an official language of Wales, jointly with English, but it has no official status in the EU. In the Welsh-language Wicipedia on the same date, there were 100,541 published articles ("List of Wikipedias" n.d.).

As these numbers include not only long articles but also very short articles called 'stubs' and articles created by bots, it is also worth looking at the depth of collaborativeness for each language on that date. Broadly, this measures how frequently articles in a given language are edited, and how much discussion takes place about the content of the articles.

**Figure 2: Comparison of the depth score of various languages**

| Depth score | Language |
|---|---|
| 901.01 | English |
| 222.09 | French |
| 205.9 | Spanish |
| 57.65 | Hungarian |
| 32.06 | Welsh |
| 30.2 | Catalan |
| 5.86 | Swedish |

Source: List of Wikipedias n.d.

As can be seen in Figure 2, Welsh therefore compared favourably with languages which have more speakers when one considers the collaboration depth score.

Collaboration is fundamental in both Wikipedia and terminology work. Unlike, for example, translation, terminology work has never been considered the enterprise of an individual and has always required a 'crowd' of sorts. Failure to be inclusive and reach a consensus regarding contentious terms rarely bodes well for the acceptance of those terms by the wider community. In the Welsh terminology work undertaken by Bangor University, over 100 editors and domain experts have collaborated over the past 25 years or so on numerous dictionaries in a wide range of fields. With a greater number of experts, so comes a greater level of expertise, as Barbara Karsch puts it in her chapter on terminology work and crowdsourcing in the Handbook of Terminology,

> The main asset of the crowd is the vast knowledge represented, access to which would not normally be open to a terminologist in her office. Nevertheless, not all input is equally valuable and a terminologist must be able to recognize good value. (Karsch 2015: 302)

It is clear here that within the discipline of terminology, a selective crowd is required. Karsch underlines this point, referring to "the selection of the crowd", "the chosen crowd" and "the right crowd" as a requirement, as opposed to the inclusion within the standardization process of an unspecified crowd of people from the general public (Karsch 2015: 302 & 291). This stems from the necessity of having a small team who know what tasks are required of them, can communicate and cooperate well together

and are capable of working efficiently and systematically following the ontological approach found in terminology standardization, compared to the ad hoc approach to article creation typical of Wikipedia.

Nevertheless, although Wikipedia's collaborative nature is perhaps less of a unique selling point than is generally perceived, there is no denying its reach and general importance as a resource. It is especially valuable for minority languages where resources may be scarce, and where there may not be many other available platforms for publishing on the web.

"An endangered language will progress if its speakers can make use of electronic technology" wrote David Crystal (Crystal 2000: 141), and it is widely acknowledged that the presence of a language on the internet and other new media is vital to its continuation and success. As a result, there is a particular enthusiasm in minority language communities for creating and freely sharing content with others as a means of language upkeep and revitalization.

For language communities where there is very little government support and public spending, such as Breton, Wikipedia offers volunteers an easy-to-use publishing and dissemination platform, and language activists can easily publish short articles on important topics in their language, in a widely accessible format and without needing special technical expertise and resources. A flourishing Wikipedia in a minority language can therefore be important to its speakers, and some minority language communities have made concerted efforts to maximise the Wikipedia content in their language. The Welsh Government has recognized Wikipedia's importance for the Welsh language online. Its 2017-21 Work Programme for the Cymraeg 2050 Strategy for the Welsh language– which aims to increase the number of Welsh speakers to one million by the year 2050 – specifically references Welsh Wicipedia, stating it will "support efforts to increase the number of Welsh-language Wicipedia pages" (Welsh Government 2017a: 35). The government has provided grant aid to agencies such as the 'Mentrau Iaith' (Language Ventures) to further this aim, and the National Library of Wales has also employed a dedicated Wikimedian in Residence – the first such in the UK – to further the same aim (Welsh Government 2017b, The National Library of Wales 2017).

As the part of this push in Wales towards creating Welsh-language Wicipedia content, Bangor University's Language Technologies Unit (LTU), home of the authors of this paper, have been asked more than once if it would be possible to share their terms and definitions with Wicipedia. As a result, the possibilities of sharing these entries with Wicipedia were explored. This was seen by the LTU team as having the potential for a mutually beneficial exchange: could Wiki content add value to the LTU's dictionaries, and could those same dictionaries add value to Wicipedia?

**Using content from the Wikipedia family of websites**

The LTU has been creating domain-specific dictionaries since 1993 (Andrews and Prys 2016), and, as a result, it is responsible for a great deal of content in the form of definitions and term entries. Many of these dictionaries were funded by external clients (mainly from the public sector in Wales), and the clients' interests needed to be considered before work that they had funded could be licensed more openly. The LTU have two large ongoing dictionary projects in the field of Education:

1. Welsh-language terminology standardization for schools and further education, entitled *Y Termiadur Addysg* (Prys and Prys 2011-2018)
2. Welsh-language terminology standardization for the eight Welsh universities teaching courses through the medium of Welsh, entitled *Geiriadur Termau'r Coleg Cymraeg Cenedlaethol* (Andrews and Prys 2010-2018).

The first terminological dictionary is funded directly by the Welsh Government and the second by the Coleg Cymraeg Cenedlaethol (Welsh National College). Initially the LTU's dictionaries were published and sold in print format and on CD-ROM. Following the rise of internet as a ubiquitous service, they are now published exclusively online and in apps, where the content is all free to access.

The terminology standardization work, including definition writing, is undertaken in an online platform called 'Maes T', which was developed in-house. The Maes T system facilitates collaboration between geographically dispersed teams of domain experts and terminologists (Andrews and Prys 2011). When dictionaries are published on this platform they are then distributed to the Welsh National Terminology Portal website and they are also distributed individually, if required, to any external websites where clients wish their own particular dictionary to be searchable (Prys, Andrews et al 2012).

All entries in the LTU's terminological dictionaries include an English term, an equivalent standardized Welsh term, and grammatical information such as part of speech and gender. Other information may also be included, such as disambiguators and definitions. The aim of including definitions is to enable the reader to recognize a concept and differentiate between it and other concepts, especially related or similar concepts. Definitions are only included when it is possible within the scope of the project; that is, when the funder commissions the LTU to do so, and time and financial resources are allocated to this task. Currently, definitions are provided in the Geiriadur Termau'r Coleg Cymraeg Cenedlaethol, which is the dictionary of terms for universities.

Online publishing has allowed the LTU to experiment with additional content in terminological entries, where before the team had been restricted to a certain extent by the size constraints of a single-volume printed dictionary and the associated printing costs. Online publishing allowed the LTU to gradually introduce additional features that would be of value to students, such as cross-references to related concepts in definitions, usage notes, mathematical equations, diagrams and images.

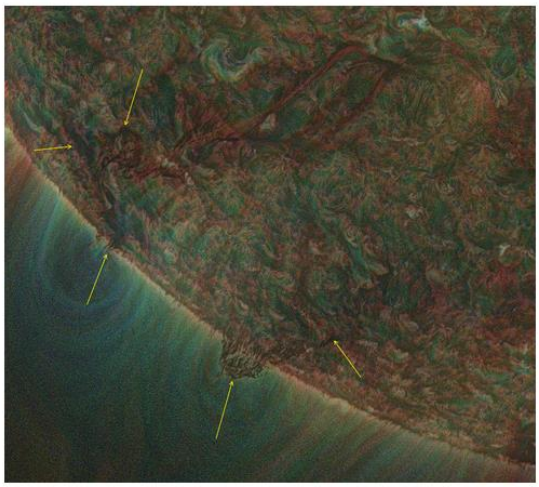
## The benefit of using Wikimedia images

The first obvious benefit of interconnecting the LTU's terminological dictionaries with Wiki-based content was access to their ready-made images which could be used to enhance terminological entries. The diagrams and images included in entries in the Geiriadur Termau'r Coleg Cymraeg Cenedlaethol in recent years had been supplied by domain experts, who were predominantly academics. These included diagrams of molecular structures and images of solar phenomena produced by the experts themselves, often as part of their research activities.

**Figure 3: Dictionary entry containing an image provided by a Welsh domain expert**

**filament** (astronomy) ffilament **eg** ffilamentau

Tafod o nwy dwys cymharol oer wedi ei ïoneiddio (~10,000K), yn gaeth mewn bwndeli cymhleth o faes magnetig yn atmosffer isel yr Haul. Mae ffilamentau'n ymddangos yn dywyll yn erbyn cryfder yr Haul y tu ôl iddynt.



A tongue of dense relatively cool ionized gas (~10,000K), held in place by complex bundles of magnetic field in the Sun's low atmosphere. Filaments appear dark against the brightness of the Sun behind them.

*Geiriadur Termau'r Coleg Cymraeg Cenedlaethol - Mathemateg a Ffiseg*

Source: Andrews and Prys 2010-2018

Not every client who commissions a terminological dictionary, however, has the time or funds to create their own images and in such cases it is possible to turn to Wikimedia Commons. Wikimedia Commons, which forms part of the Wikipedia family of websites, is a repository of over 40 million media files ("Commons: Welcome" n.d.) that are licensed under public domain and various free licences with different degrees of permissibility.

The first dictionary where the LTU incorporated Wikimedia files was a short dictionary entitled *Buchod Cwta / Ladybirds* (Brown, Elias et al 2014) which contains names for the ladybird family *Coccinelidae* and which is hosted in the Maes T platform for Cymdeithas Edward Llwyd, an association of Welsh naturalists. The Wikimedia files included a large number of high quality photographs of various species. The inclusion of these images in the ladybird dictionary was seen by Cymdeithas Edward Llwyd as being vital to enabling its users to recognize the different species of the ladybird family, given that the dictionary did not include textual definitions.

**Figure 4: Dictionary entry containing a Wikimedia image**


Source: Brown, Elias, et al 2014
Image by Pudding4brains [Public domain], from Wikimedia Commons

## Licensing

For a photograph to be included in the dictionary, however, its licence had to be examined, to ensure that the LTU's intended use was acceptable under the terms of that particular licence. There are a number of common licences which are referred to as being 'free', and these vary in terms of their permissibility, with some placing more restrictions than others on the reuse of the content. For the dictionary of ladybird names, photographs published with one of the three following licences were deemed appropriate for use: Public Domain, Creative Commons Attribution, known as CC BY, and Creative Commons Attribution-ShareAlike, known as CC BY-SA.

Public domain images are free for use without restriction. CC BY may be used without restriction as long as the author of the image is acknowledged. CC BY-SA may be used without restriction as long as the author is acknowledged and the user of the image also agrees to share their content with others. The ShareAlike clause can cause problems for content which is reused from other sources where different pre-existing licences are in force, as is discussed later. Cymdeithas Edward Llwyd, however, agreed to share the dictionary with others under the ShareAlike licence and to transfer its dictionaries to Welsh-language Wicipedia for distribution under the terms of CC BY-SA.

## Methodology

Two methods have been used by the LTU to incorporate Wikimedia images in dictionaries. As the dictionary of ladybird names contained fewer than 50 names, it was possible for the members of Cymdeithas Edward Llwyd to manually search Wikimedia for suitable images with an appropriate licence, then add an image URL to the relevant entry in the Excel file in which the dictionary had been created. This file was then imported into the Maes T platform and published as an online dictionary. The licensing and attribution information for each image was included in the URL provided in Wikimedia and this could be seen in rollover text on every image in the online dictionary.

Some months after this dictionary was published, Cymdeithas Edward Llwyd requested that the same work should be done on a much larger dictionary of over 9,500 names of birds, entitled *Adar y Byd/ Birds of the World* (Fear, Elias et al 2015). It became clear that an automated solution would be required, since searching for these images by hand would be very laborious. Therefore, the LTU created a code which could map the dictionary entries to corresponding Creative Commons photographs from Wikimedia Commons and import them into the relevant entries in the Maes T system. The code created would search the API service of the Wikidata website to find an image file corresponding to the scientific Latin name of the animal, or else corresponding to its Welsh or English common name.

**Figure 5: Sample of the code used to find relevant Wiki media files for inclusion in the dictionary**

```
SELECT ?item ?file
{
  {
    { ?item wdt:P225 "Cygnus columbianus bewickii"}
  UNION
    { ?item wdt:P225 "cygnus columbianus bewickii"}
  UNION
    { ?item rdfs:label "Alarch Bewick"@cy}
  UNION
    { ?item rdfs:label "Bewick's swan"@en}
} .
?item wdt:P18 ?file .filter ( bound ( ?file ) )
}
```

Wikidata, which is another website within the Wikipedia family, is a database of linked data, indexed by concept. It does for data what Wikimedia Commons does for media files. It is structured to be read and edited by both humans and machines, and is consequently more suitable for an automated solution. A code similar to the sample in Figure 5 above would be sent automatically to https://query.wikidata.org/sparql. If more than one image was found in Wikidata, the first of these would be included in the dictionary. A second code would then create a HTML string for inclusion in the dictionary, which used what was at the time an experimental API for extracting the attribution of Wikimedia artefacts, so that the dictionary entries would comply with the licensing requirements. This second code was created in the absence of a mechanism in Wikimedia itself for summarizing licensing and attribution information for the inclusion of artefacts into third-party collections such as dictionaries. This summarized information would then be visible in rollover text on every image in the online dictionary, as seen in Figure 6 below. A third piece of code would do two further things. First, it would search Wicipedia using the Welsh common name to see if the encyclopaedia included an article in Welsh about the bird. If such an article existed, the code created a link in the dictionary to the external article. Secondly, it would search for a sound file relating to the bird on Wikidata and Wikimedia Commons and then include this also in the dictionary.

**Figure 6: Dictionary entry containing a Wikimedia image, sound file and link to a relevant external Wikipedia article**


Source: Fear, Elias et al 2015
Image by Andreas Trepte [CC BY-SA 2.5 (https://creativecommons.org/licenses/by-sa/2.5)], from Wikimedia Commons

Use of this code was very successful and subsequently some 8,000 entries included a photograph. Given its success, the same method was then used for adding images to a third dictionary, *Gwyfynod, Glöynnod Byw a Gweision Neidr / Moths, Butterflies and Dragonflie*s (Brown, Elias et al 2009).

It was discovered, however, that results were more likely to be accurate if the search for Wikimedia images was carried out using the Latin name of the animal rather than a common name. The vast majority of entries were similar to Figure 6 above, although Figure 7 shows one of the potential pitfalls of searching for images using the common name.

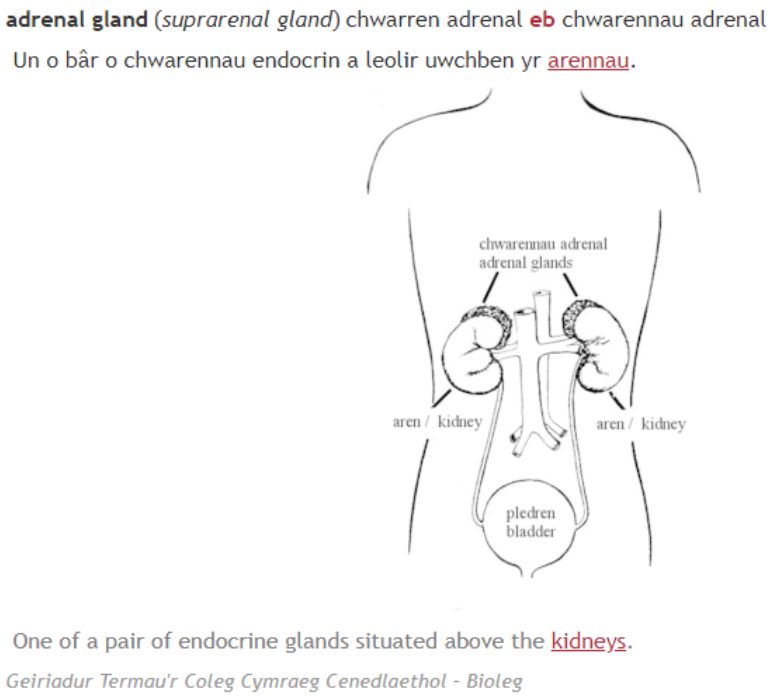
**Figure 7: Sample of the problems which can arise from searching Wikimedia images using the common name of an animal**


Source: Brown, Elias et al 2009
Image by Steve Collis from Melbourne, Australia (Doctor Who Experience) [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0)], from Wikimedia Commons

Here, *Eriopygodes imbecilla*, a moth whose common name is *silurian*, was represented by an image of another concept, namely a character from a fictional race called the *Silurians* who appear in the *Doctor Who* science fiction television series. Use of the Latin name in automatic searching avoids this type of issue.

For the Geiriadur Termau'r Coleg Cymraeg Cenedlaethol dictionary of terms for universities, the LTU reverted to inputting Wiki content by hand, because including images might not have been appropriate in many term entries. In fields such as earth sciences and biology, images often added value to certain term entries, and the ability to adapt images to include Welsh terms in labels also proved useful, such as in Figure 8 below.

**Figure 8: Biology term entry with Wiki image**




Source: Andrews and Prys 2010-2018
Image by Pearson Scott Foresman [Public domain], from Wikimedia Commons

## Repercussions of sharing content

The LTU concluded that, Doctor Who-type pitfalls notwithstanding, the addition of Wiki content to terminological dictionaries gave added value to the dictionaries. On the other side of this relationship between the Wiki world and terminological dictionaries, however, the effects on a standardized term list of releasing the names of animals and attendant information for free use to Wicipedia appeared more problematic. Of 1042 names of moths and butterflies given in a list format in a Welsh Wicipedia article ("Rhestr gwyfynod a gloÿnnod byw" n.d.), which references Cymdeithas Edward Llwyd as a source, 171 Welsh names do not correspond to the name standardized by this group of experts in their dictionary published in 2009. That is over 16%. Examples are included in Figure 9.

**Figure 9: Non-corresponding names**

| No. | Scientific Latin name | Name in the dictionary | Name in the Wicipedia article |
|---|---|---|---|
| 1 | *Trichiocercus sparshalli* | sidan cynffonnog | sidan cynffonog |
| 2 | *Deilephila elpenor* | gwalch-wyfyn helyglys | gwalchwyfyn yr helyglys |
| 3 | *Parnassius phoebus* | glöyn apolo bach | glo dau-ddot yn Apolo bach |
| 4 | *Hada plebeja* | pali siswrn | arennau disglair |
| 5 | *Periphanes delphinii* | gwyfyn melynwellt porffor | melyngoch porffor |
| 6 | *Fagivorina arenaria* | rhisglyn smotiog | rhisglyn brych |
| 7 | *Alcis repandata* | rhisglyn brych | rhisglyn brych |

In 1-3, the common names in the Wicipedia article have variant or non-standard spellings compared to those in the dictionary, and it seems that reproducing the accented character ö is also problematic. In 4-5 the Wicipedia article uses completely different common names from those in the dictionary. In 6-7, the dictionary gives two different names to two different species, whereas Wicipedia uses the same name for both.

This may not be so important in certain general-language environments, but the nomenclature and classification of species as described in the scientific literature is a specialist, technical domain. When the LTU team examined the article-naming practices in Wikipedia, they saw that the Wiki goals include

“recognisability” which is defined as “a name that someone familiar with, although not necessarily an expert in, the subject area will recognize” (“Wikipedia: How2title” n.d.). In the case of the names of species it is very possible that the more recognizable name for the lay person may be a less formal, non-technical name. In the Welsh Wicipedia naming goals they note:

Yr hyn sy'n bwysicach na'r dewis o derm yw cofio rhoi'r termau amgen yn yr erthygl a chreu tudalennau ailgyfeirio o'r term amgen. (“Wicipedia: Arddull” n.d.)

Translation: What is more important than the choice of term is remembering to include alternative terms in the article and creating redirect links from the alternative term.

The use of the phrase “choice of term” coupled with a desire to include all synonyms shows the fundamental difference in practice between the naming guidelines of Wikipedia and of terminological dictionaries, where only the prescribed form is given.

The repercussions of sharing a list of standardized terms or names with Wicipedia – and Wicipedia contributors then failing to include an indication of the standardized status of specific names – are of concern to terminologists and should also be of concern to domain experts. What was once a standardized list becomes a mixture of standardized and non-standardized names, with no indication of their relative status, as is the case now with the list of moths and butterflies. Their status is not clear to the reader unless he or she cross-references specific terms using both sources. Therefore, should a group involved in standardizing terminology be facilitating the fragmentation and editing of its term lists by the general public? The terminologists at the LTU are hesitant about adopting such an approach.

As the LTU had, however, benefited from the use of Wikimedia Commons and Wikidata, the team wanted to explore how they could best reciprocate with the Wikipedia family and contribute to its content. Given the worrying repercussions of releasing lists of standardized terms from the protective environment of a non-editable dictionary to an editable platform with different naming policies, it was felt that releasing definitions might be a better contribution. The obvious dictionary from which to do that was Geiriadur Termau’r Coleg Cymraeg Cenedlaethol, since it includes scholarly definitions, and has also itself benefited from incorporating Wikimedia Commons images. As the LTU team explored Welsh Wicipedia, they discovered that several Wicipedia articles included definitions which had been copied and pasted from this dictionary, or exclusively consisted of these definitions – in breach of the dictionary’s copyright. Since bringing this to the attention of Wicipedia, these definitions have been adapted to avoid the licensing problem. The LTU understood from this discovery that contributors to Wicipedia already considered this dictionary’s definitions useful starting points for articles, and that releasing these would benefit Wicipedia and lead to the creation of more Welsh content. The definitions copied and pasted were from earth sciences and biology. Therefore, the LTU explored the sciences as a promising starting point for their investigations into the possibilities which might be open to them for sharing content. The LTU also resolved to improve the notices on their website about licensing and copyright.

**Releasing definitions to Welsh Wicipedia**

The first step was to compare natural science entries from Geiriadur Termau’r Coleg Cymraeg Cenedlaethol with Wicipedia articles to see how many of the dictionary headwords did not have a corresponding Welsh article on Wicipedia. Of 1409 forestry entries, for example, only 93 had a Wicipedia entry, and of 272 biology entries, only 39 had a corresponding Welsh Wicipedia entry.

Even accounting for the type of headword that might not be appropriate as the headword of an encyclopaedic article, such as related word forms – an adjective, noun and verb – where perhaps only the noun might be an article topic, it appeared that the above-mentioned dictionary of terms for universities might help fill some blanks in Welsh Wicipedia content.

The LTU team soon encountered a hurdle however, namely, how to license the definitions. The university dictionary is currently divided into 15 domains and while it appears to the user as one dictionary, in the back end many of the terms for each domain originated in discrete dictionaries, each one with different licensing conditions. Some of these are works which the LTU were given permission to incorporate in this dictionary, subject to specific permissions. Domain-specific dictionaries incorporated in the dictionary include:

- three print dictionaries in the fields of forestry, law and psychology, digitized, revised and incorporated with the permission of its copyright holder, Bangor University, but including legacy data subject to prior licensing agreements;
- one French-Canadian biology dictionary adapted for Welsh by the LTU and incorporated with permission of its copyright holder, subject to certain restrictions;
- one geology dictionary by an individual expert, with permission to incorporate it into the LTU dictionary, but not to use it in further derivative works;
- others developed at Bangor University in collaboration with domain experts across Wales.

As there were so many sources included in the dictionary the team concluded that was impossible to license it in its entirety under one type of license. They therefore decided to experiment on a very small scale by licensing five psychology definitions – where there were no pre-existing licensing restrictions – as CC BY.

In order to be able to add new entries to Wicipedia, it was necessary to sign up as a contributor and make 10 edits to existing articles. Then the LTU began 5 new articles by following a link in an existing article to an as-yet unwritten article with an existing title and adding the definition from the dictionary and a reference to the dictionary in a footnote.

**Incompatibilities between terminological definitions and Wikipedia articles**

Fundamental incompatibilities between terminological definitions and encyclopaedic descriptions became apparent when the Wicipedia team informed the LTU that the submitted definitions did not meet the requirements of a Wicipedia article. Of these, the lack of outgoing hyperlinks and category information could be easily overcome, but the remaining issues, namely lack of encyclopaedic-level depth, lack of citations and shortage of references, were more profound. The LTU were advised that the quickest way to expand an entry was to translate the first paragraphs of the corresponding entry in another language to Welsh. This did not correspond with the LTU's aim to share its original Welsh definitions.

Within less than 12 hours, of the five entries created:

- one entry was rewritten by a Wicipedia editor, as an example to show how Wicipedia entries should be written should the LTU want to continue to contribute and comply with Wicipedia rules. This included the removal of the supplied dictionary definition.
- one entry had a warning applied to the article stating that it did not reach the necessary standard to be included in Wicipedia, and that it could be deleted within a week if it was not improved.
- three entries were unchanged, including one single-sentence article.

The LTU's intention was not to write or translate encyclopaedic-length entries in Wicipedia, but rather to share its definitions as a starting point for others to develop these into more in-depth encyclopaedic entries. Wikipedia notes that while a definition "may be enough to qualify as a stub, Wikipedia is not a dictionary" (Wikipedia: Stub. n.d.). Stubs are articles which are too short to provide encyclopaedic coverage but which can be expanded. The LTU continues to believe that its entries could serve as appropriate stubs to be expanded upon by others at a later date. The LTU also considered contributing term lists, much as Cymdeithas Edward Llwyd had done. Term lists within single articles are found throughout Wikipedia in different languages (for example "Rhestr pysgod" n.d., "Rhestr adar" n.d., "Roll laboused" n.d. and "Llista de minerals" n.d.). However, this would mean ceding editorial control

of the term list's content to Wicipedia editors. As editorial control is an intrinsic part of term standardization it was decided that this would not a suitable course of action.

## Conclusion

The LTU's experimentation with contributing dictionary entries to Welsh Wicipedia has served to highlight the need for terminological dictionaries and crowdsourced encyclopaedias to remain as separate and distinct resources with differing aims, whilst identifying areas where content can be shared between both types of resources. The LTU's terminological dictionaries have benefited enormously from the addition of images taken from Wikimedia Commons and Wikidata, although the team has had to pay close attention to the licensing terms so that the LTU does not transgress any pre-existing licensing conditions. Our experiences of sharing entries with Wicipedia were therefore somewhat mixed. Whilst there is merit in the idea of providing terminological definitions as a starting point for encyclopaedic articles, the success of such an endeavour would be reliant on Wicipedia administrators' interpretation of their own guidelines. Without the resources to expand concise terminological dictionaries into encyclopaedic descriptions (something that is not within the scope of the LTU's current terminological dictionaries), the risk remains that any terminological entries shared with Wicipedia will be rejected on the grounds that they are too concise to form an encyclopaedic article. The LTU has also realized that it needs to take a more granular approach to the copyright and licensing of its dictionaries and clarify this on the website so that there is absolute clarity in regard to what can legitimately be reused and repurposed by others.

The role of the terminology arm of the LTU is to create and disseminate standardized Welsh terms, rather than produce encyclopaedic entries, and this must remain its primary responsibility. The greatest concern of the team, as standardizers of terminology, is that the standardized terms are recognized and adhered to. As a result of this exercise, the LTU and Wicipedia are discussing ways of denoting a standardized term and linking it to its authoritative source. From a terminology standardization perspective, the inclusion of synonyms and colloquial or dialectical forms could be acceptable as long as the standardized form is clearly marked. This would make such entries more compatible with the form encouraged in an encyclopaedic entry, where there is much more scope for elaboration. It is envisaged that a template assigning an 'authority ID' to standardized terms and definitions would be used to facilitate this functionality, especially if the process of applying such a feature is to be automated. However, safeguards will also need to be put in place to ensure long-term consistency should changes be made to the article or the standardized term in the future.

Should a process for identifying standardized terms on Wicipedia using an 'authority ID' be established, this would allow the LTU to share suitable terms and definitions as stubs for Wicipedia entries. This would provide a useful basis for Wicipedia's community of volunteers to extend and build upon, secure in the knowledge that the identity of the standardized term and definition is acknowledged and referenced. In this way the environments of both the terminological dictionaries and Welsh Wicipedia would be enriched, and both would benefit from enhanced collaboration. Discussions to put such a process in place are underway.

## References

Andrews, Tegau and Gruffudd Prys. (2011). "The Maes T System and its use in the Welsh-Medium Higher Education Terminology Project." In *Proceedings of CHAT 2011: Creation, Harmonization and Application of Terminology Resources*, Riga, Latvia, 11 May 2011; Gornostay, T., Vasil ̧jevs, A., Eds.; NEALT: Tartu, Estonia. 49–50.

Andrews, Tegau and Gruffudd Prys. (2016). "Terminology Standardization in Education and the Construction of Resources: The Welsh Experience." *Education Sciences 6*. Retrieved 14 May 2018, from http://www.mdpi.com/2227-7102/6/1/2/htm.

Andrews, Tegau and Delyth Prys. (2010-2018). *Geiriadur Termau'r Coleg Cymraeg Cenedlaethol*. Retrieved 10 May 2018, from http://termau.cymru/.

British Standards Institution. (2001). *Project management guidelines for terminology standardization*. (BS ISO 15188: 2001).

British Standards Institution. (2009). *Terminology work – Principles and methods.* (BS ISO 704: 2009).

Brown, Duncan, Twm Elias, Bruce Griffiths, Huw John Huws, and Dafydd Lewis. (2009). *Gwyfynod, Glöynnod Byw a Gweision Neidr / Moths, Butterflies and Dragonflies*. Cymdeithas Edward Llwyd. Retrieved 01 November 2017, from http://termau.cymru/.

Brown, Duncan, Twm Elias, Bruce Griffiths and Selwyn Williams. (2014). *Buchod Cwta / Ladybirds*. Cymdeithas Edward Llwyd. Retrieved 08 November 2017, from http://termau.cymru/.

Crystal, David (2000). *Language Death*. Cambridge: Cambridge University Press. 141.

Fear, Davyth, Twm Elias, Eilir Evans, Bruce Griffiths and Duncan Brown. (2015). *Adar Y Byd / Birds of the World*. Cymdeithas Edward Llwyd. Retrieved 08 November 2017, from http://termau.cymru/.

Karsch, Barbara. (2015). 'Terminology work and crowdsourcing: Coming to terms with the crowd'. In Hendrik J. Kockaert and Frieda Steurs (eds.). *Handbook of terminology: 1*. Amsterdam [u.a.]: Benjamins. 291-302.

List of Wikipedias. (n.d.). In *Wikipedia, The Free Encyclopedia*. Retrieved 14 May 2018, from https://en.wikipedia.org/wiki/List_of_Wikipedias.

Llista de minerals. (n.d.). In *Viquipèdia l'enciclopèdia lliure*. Retrieved on 18 May 2018, from https://ca.wikipedia.org/wiki/Llista_de_minerals.

Office for National Statistics. (2012). *2011 Census: Key Statistics for Wales, March 2011*. Retrieved on 17 May 2018, from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuskeystatisticsforwales/2012-12-11.

Prys, Delyth and Gruffudd Prys. (2011-2018). *Y Termiadur Addysg*. Retrieved 08 November 2017, from http://termau.cymru/.

Prys, Gruffudd, Tegau Andrews, Dewi B. Jones and Delyth Prys. (2012). "Distributing terminology resources online: multiple outlet and centralized outlet distribution models in Wales." In *Proceedings of CHAT 2012: Creation, Harmonization and Application of Terminology Resources*, Madrid, Spain, 22 June 2012; Gornostay, T., Ed.; Linköping University Electronic Press: Linköping, Sweden. 37–40. Retrieved 16 May 2018, from http://www.ep.liu.se/ecp/072/005/ecp12072005.pdf.

Rhestr adar Prydain. (n.d.). In *Wicipedia: Y Gwyddoniadur Rhydd*. Retrieved 17 May 2018, from https://cy.wikipedia.org/wiki/Rhestr_adar_Prydain

Rhestr gwyfynod a gloÿnnod byw. (n.d.). In *Wicipedia: Y Gwyddoniadur Rhydd*. Retrieved 08 November 2017, from https://cy.wikipedia.org/wiki/Rhestr_gwyfynod_a_glo%C3%BFnnod_byw.

Rhestr pysgod, molysgiaid, cramenogion ayyb. (n.d.). In *Wicipedia: Y Gwyddoniadur Rhydd*. Retrieved 17 May 2018, from https://cy.wikipedia.org/wiki/Rhestr_pysgod,_molysgiaid,_cramenogion_ayyb.

Roll laboused Breizh. (n.d.). In *Wikipedia, An hollouezíadur digor*. Retrieved 17 May 2018, from https://br.wikipedia.org/wiki/Roll_laboused_Breizh.

Safer, Morley. (2015). "Wikimania." *CBS News*. Retrieved 14 May 2018 from https://www.cbsnews.com/news/wikipedia-jimmy-wales-morley-safer-60-minutes/.

The National Library of Wales. (2017). "*The National Library of Wales appoint UK's first permanent Wikimedian.* 2017 Press releases. Retrieved on 16 May 2018, from https://www.library.wales/information-for/press-and-media/press-releases/2017-press-releases/the-national-library-of-wales-appoint-uks-first-permanent-wikimedian/.

Welsh Government. (2017a). *Cymraeg 2050: A million Welsh speakers – Work programme 2017-21*. Retrieved 10 May 2018, from https://gov.wales/docs/dcells/publications/170711-cymraeg-2050-work-programme-eng-v2.pdf.

Welsh Government. (2017b). *Projects which Get Creative with Cymraeg announced.* Retrieved 16 May 2018, from https://gov.wales/newsroom/welshlanguage/2017/projects-which-get-creative-with-cymraeg-announced/?skip=1&lang=en

Wicipedia: Arddull. (n.d.). In *Wicipedia: Y Gwyddoniadur Rhydd*. Retrieved 08 November 2017, from https://cy.wikipedia.org/wiki/Wicipedia:Arddull.

Wikipedia: The perfect article. (n.d.). In *Wikipedia, The Free Encyclopedia*. Retrieved 08 November 2017, from https://en.wikipedia.org/wiki/Wikipedia:The_perfect_article.

Wikipedia: Introduction. (n.d.). In *Wikipedia, The Free Encyclopedia.* Retrieved 08 November 2017, from https://en.wikipedia.org/wiki/Wikipedia:Introduction.

Wikipedia: How2title. (n.d.). In *Wikipedia, The Free Encyclopedia.* Retrieved 08 November 2017, from https://en.wikipedia.org/wiki/Wikipedia:How2title.

Wikipedia: Stub. (n.d.). In *Wikipedia, The Free Encyclopedia*. Retrieved 08 November 2017, from https://en.wikipedia.org/wiki/Wikipedia:Stub.